

# Med-Flamingo: a Multimodal Medical Few-shot Learner

Year: 2023 | Citations: 404 | Authors: Michael Moor, Qian Huang, Shirley Wu, Michihiro Yasunaga, Cyril Zakka

---

## Abstract

Medicine, by its nature, is a multifaceted domain that requires the synthesis of information across various modalities. Medical generative vision-language models (VLMs) make a first step in this direction and promise many exciting clinical applications. However, existing models typically have to be fine-tuned on sizeable down-stream datasets, which poses a significant limitation as in many medical applications data is scarce, necessitating models that are capable of learning from few examples in real-time. Here we propose Med-Flamingo, a multimodal few-shot learner adapted to the medical domain. Based on OpenFlamingo-9B, we continue pre-training on paired and interleaved medical image-text data from publications and textbooks. Med-Flamingo unlocks few-shot generative medical visual question answering (VQA) abilities, which we evaluate on several datasets including a novel challenging open-ended VQA dataset of visual USMLE-style problems. Furthermore, we conduct the first human evaluation for generative medical VQA where physicians review the problems and blinded generations in an interactive app. Med-Flamingo improves performance in generative medical VQA by up to 20% in clinician's rating and firstly enables multimodal medical few-shot adaptations, such as rationale generation. We release our model, code, and evaluation app under <https://github.com/snap-stanford/med-flamingo>.